AFRL-RH-WP-TR-2010-0065

Predictive Validity of the Air Force Officer Qualifying Test (AFOQT) for Non-Rated Officer Specialties

Thomas R. Carretta
Warfighter Interface Division
Supervisory Control Interfaces Branch
Wright-Patterson AFB OH 45433-7511

March 2009

Interim Report for November 2008 to March 2009

Air Force Research Laboratory
711th Human Performance Wing
Human Effectiveness Directorate
Warfighter Interface Division
Supervisory Control Interfaces Branch
Wright-Patterson AFB, OH 45433

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88th ABW Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RH-WP-TR-2010-0065 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signature//

Thomas R. Carretta, Ph.D. Work Unit Manager
Supervisory Control Interfaces Branch

//signature//

Michael A. Stropki, Chief
Warfighter Interface Division
Human Effectiveness Directorate
711th Human Performance Wing
Air Force Research Laboratory

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved

OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
12-03-2009	Interim	November 2008 – March 2009

4. TITLE AND SUBTITLE
Predictive Validity of the Air Force Officer Qualifying Test (AFOQT) for Non-Rated Officer Specialties

5a. CONTRACT NUMBER
FA8650-08-D-6801-0013

5b. GRANT NUMBER
NA

5c. PROGRAM ELEMENT NUMBER
62202F

6. AUTHOR(S
Thomas R. Carretta

5d. PROJECT NUMBER
2313

5e. TASK NUMBER
HC

5f. WORK UNIT NUMBER
2313HC58

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Materiel Command
Air Force Research Laboratory
711th Human Performance Wing
Human Effectiveness Directorate
Warfighter Interface Division
Supervisory Control Interfaces Branch
Wright-Patterson AFB OH 45433

10. SPONSOR/MONITOR'S ACRONYM(S)
711 HPW/RHCI

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-RH-WP-TR-2010-0065

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; Distribution is unlimited.

13. SUPPLEMENTARY NOTES
88 ABW Cleared 04/06/2009; 88ABW-09-1347.

14. ABSTRACT
A meta-analysis was performed to examine the generalizability of the predictive validity of the Air Force Officer Qualifying Test (AFOQT) operational composites against technical training performance for 14 non-rated Air Force Specialties (AFSs). AFOQT data were from Form Q and were used to compute composites based on Form S specifications. All five operational composites (Verbal, Quantitative, Academic Aptitude, Pilot, and Navigator/Technical) were included in the analyses. The criterion was technical training final grade. Analyses began by examination of the observed correlation between the AFOQT composites and technical training final grade for each officer training course. The meta-analysis of the observed correlations was corrected only for sampling error. The observed correlations then were corrected for range restriction using the multivariate method (Lawley, 1943) and the meta-analysis was repeated. The range-restriction corrected correlations were then corrected for unreliability (Hunter & Schmidt, 2004) of the test scores and training criterion and the meta-analysis was repeated.

15. SUBJECT TERMS
Air Force Officer Qualifying Test, AFOQT, predictive validity, officer technical training

16. SECURITY CLASSIFICATION OF: Unclassified			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Thomas R. Carretta
U	U	U	SAR	23	19b. TELEPONE NUMBER (*Include area code)*

This page is intentionally left blank.

ABSTRACT

A meta-analysis was performed to examine the generalizability of the predictive validity of the Air Force Officer Qualifying Test (AFOQT) operational composites against technical training performance for 14 non-rated Air Force Specialties (AFSs). AFOQT data were from Form Q and were used to compute composites based on Form S specifications. All five operational composites (Verbal, Quantitative, Academic Aptitude, Pilot, and Navigator/Technical) were included in the analyses. The criterion was technical training final grade. Analyses began by examination of the observed correlation between the AFOQT composites and technical training final grade for each officer training course. The meta-analysis of the observed correlations was corrected only for sampling error. The observed correlations then were corrected for range restriction using the multivariate method (Lawley, 1943) and the meta-analysis was repeated. The range-restriction corrected correlations were then corrected for unreliability (Hunter & Schmidt, 2004) of the test scores and training criterion and the meta-analysis was repeated. This third set of correlations provides a theoretical estimate of the predictiveness of the composites when perfectly reliable measures are available. Ninety percent (63 out of 70) of the observed correlations between the AFOQT composites and average officer technical training grades were statistically significant at or beyond the .05 level, thus supporting its general value in selection of non-rated officers. Results of the meta-analyses of the observed correlations indicated that the predictive validity for four of the five AFOQT composites (Verbal, Academic Aptitude, Pilot, and Navigator/ Technical) was not the same across the non-rated officer training specialties in the current study. Validity generalization was observed for only the Quantitative composite. After correction for range restriction, four of the five composites demonstrated validity generalization across the training specialties. This means that the true validity of the AFOQT composites (with the exception of Verbal), was consistent across the officer training specialties. The mean validity coefficients for the Quantitative (.3499), Academic Aptitude (.3878), Pilot (.3525), and Navigator/Technical (.3796) composites are the best estimates of the average validity across all officer specialties. Additional efforts to examine the generalizability of the validity of AFOQT composites should expand to include a broader range of occupational specialties. Expanding the breadth of training specialties would allow the potential moderating effect of occupational similarity to be examined.

This page intentionally left blank.

CONTENTS

	Page
PREFACE ……..…….	**vii**
1.0 INTRODUCTION …….	**1**
2.0 METHOD ……..	**2**
2.1 Participants ……..	2
2.2 Measures …….	3
2.3 Analyses ……..……...……	5
3.0 RESULTS …….……	**6**
3.1 Observed Correlations ……...……	6
3.2 Range-Restriction Corrected Correlations ……..	7
3.3 Range Restriction and Reliability Corrected Correlations ……..	9
4.0 DISCUSSION …….……	**11**
**BIBLIOGRAPHY....……..…….. **	**12**

TABLES

No.	Title	Page
1	Composition of AFOQT Form S Aptitude Composites …….....	3
2	Correlations between AFOQT Composites and Officer Technical Training Grades: Observed ……...	6
3	Correlations between AFOQT Composites and Officer Technical Training Grades: Range-Restriction Corrected ……...	8
4	Correlations between AFOQT Composites and Officer Technical Training Grades: Range Restriction and Reliability Corrected ……	10

GLOSSARY

AF/A1PF	Air Force, Force Management Policy Division
AFOQT	Air Force Office Qualifying Test
AFPC	Air Force Personnel Center
AI	Aviation Information
AR	Arithmetic Reasoning
BC	Block Counting
GS	General Science
HF	Hidden Figures
HRRD	Human Resources Data Bank
IC	Istrument Comprehension
MK	Math Knowledge
OTS	Officer Training School
RB	Rotated Blocks
ROTC	Reserve Officer Training Corps
TR	Table Reading
VA	Verbal Analogies
WK	Word Knowledge

PREFACE

This report describes activities performed in support of USAF personnel selection and classification (AF/A1PF), Work Unit 2313HC58. The author thanks Mr. Ken Schwartz (AETC/DPSF) and the Air Force Personnel Center (AFPC) Human Resources Research Data Bank (HRRD) for support in the development of the database used in this study.

This page intentionally left blank.

1.0 INTRODUCTION

The Air Force Office Qualifying Test (AFOQT; Carretta & Ree, 1996) is used to award U.S. Air Force Reserve Officer Training Corps (ROTC) scholarships and to qualify applicants for officer commissioning through the ROTC and Officer Training School (OTS) programs. The AFOQT also is used to qualify applicants for aircrew training as pilots, combat system operators (formerly navigators), and air battle managers. The AFOQT has been validated against officer training performance (Roberts & Skinner, 1996), several aircrew training performance criteria including passing/failing training, training grades, and class rank (Carretta, in press; Carretta & Ree, 1995a, 2003; Olea & Ree, 1994), and several non-rated officer jobs (Arth, 1986; Arth & Skinner, 1986; Finegold & Rogers, 1985; Hartke & Short, 1988).

The current form of the AFOQT (Form S) was operationally implemented in 2005 and consists of 11 cognitive subtests and an experimental personality inventory. For operational use, the cognitive subtests are combined into five overlapping composites as shown in Table 1. The Verbal, Quantitative, and Academic Aptitude composites are used to qualify applicants for ROTC and OTS officer commissioning programs. The Pilot and Navigator/Technical composites are used to qualify applicants for aircrew training. Air Force Instruction 36-2013 (United States Air Force, 2006) provides AFOQT minimum qualifying score requirements for officer commissioning and aircrew training. The minimum qualifying scores for officer commissioning are at least the 15th percentile on the Verbal composite and at least the 10th percentile on the Quantitative composite. ROTC and OTS aircrew training applicants must first qualify for officer commissioning meeting minimum requirements for the AFOQT Verbal and Quantitative composites. In addition, they must meet minimum qualifying scores for the Pilot and Navigator/Technical composites. The minimum qualifying scores for aircrew training vary by program, commissioning source, and for pilot training whether the applicant has a private pilot's certificate. For many non-rated officer training specialties[1], no additional AFOQT requirements exist other than the minimum requirements for officer commissioning.

[1] Applicants for some non-rated officer training specialties (e.g., medical doctors, dentists, legal) do not require qualification on the basis of AFOQT scores. They are referred to as non-line officers. Non-line officer specialties require appropriate college degrees and training. Upon entry into the Air Force, non-line officers complete an abbreviated officer training course.

Since the implementation of AFOQT Form O in 1981, the development and implementation cycle for new forms has been about seven or eight years. AFOQT Form S was implemented in June 2005. Two lines of research are underway during the current AFOQT development cycle. The first is focused on development of content specifications for Form T. As part of this effort, focus groups are being conducted with Air Force officers in rated and non-rated career fields to identify critical knowledge, skills, abilities, and other characteristics (KSAOs) for Air Force officer and technical training programs. Responses from the focus groups will be used to develop on-line occupational surveys that will be administered to approximately 10,000 Air Force officers to determine the importance of the KSAOs to career success. The results will be used to guide the identification of constructs to supplement existing AFOQT content. The second line of research is focused on the evaluation of the predictive validity of AFOQT Form S versus training performance (e.g., Carretta, in press). To this end, the current study examined the predictive validity of the AFOQT composites versus training performance in several non-rated officer specialties. Results will provide a baseline of the predictive utility of the AFOQT for non-rated specialties.

2.0 METHOD

2.1 Participants

The sample consisted of 10,542 USAF officers who had tested on the AFOQT Form Q and subsequently attended one of 14 technical training courses. The training courses were Combat Control (13D1AB), Airfield Operations (13M1), Space and Missile (13S1), Space and Missile follow-on (13S1X), Intelligence (14N1), Weather (15W1), Aircraft Maintenance (21A1), Munitions and Munitions Maintenance – Conventional (21M1C), Munitions and Munitions Maintenance – Non-conventional (21M1NC), Logistics Readiness (21R1), Security Forces (31P1), Communications-Information Systems (33S1), Communications Officer Engineering (33S3A), and Manpower and Personnel (37F1). Sample sizes ranged from 16 (Combat Control) to 2,190 (Communications-Information Systems) with an average sample size of 753 students. The criterion was final technical training course grade which is based on several written tests and ranged from 70 to 100.

2.2 Measures

Participants tested on AFOQT Form Q, which consisted of 16 cognitive subtests. When AFOQT Form S was implemented in July 2005, five of the subtests from previous forms (O, P, and Q) had been removed. AFOQT Form S consists of 11 cognitive subtests that are combined into five composites (see Table 1). For the purpose of this study, AFOQT raw score composites were computed on the basis of the Form S content and composite specifications. Personnel decisions including qualification for officer commissioning programs and aircrew training are made, in part, on the basis of the composites. Brief descriptions of the AFOQT subtests grouped by content are presented below.

Table 1. Composition of AFOQT Form S Aptitude Composites

	Composite				
Subtest	Verbal (V)	Quantitative (Q)	Academic Aptitude (AA)	Pilot (P)	Navigator/ Technical (N/T)
Verbal Analogies (VA)	X		X		X
Arithmetic Reasoning (AR)		X	X	X	X
Word Knowledge (WK)	X		X		
Math Knowledge (MK)		X	X	X	X
Instrument Comprehension (IC)				X	
Block Counting (BC)					X
Table Reading (TR)				X	X
Aviation Information (AI)				X	
Rotated Blocks (RB)					
General Science (GS)					X
Hidden Figures (HF)					
Self-Description Inventory (SDI+)					

Note. Although RB and HF were retained in AFOQT Form S, they do not contribute to any of the operational composites. The SDI+ is an experimental non-cognitive subtest.

Confirmatory factor analyses of the AFOQT Form S subtests have shown it to measure general intelligence and the five content-specific factors of verbal, quantitative, spatial, aviation

knowledge, and processing speed (Drasgow, Nye, Carretta, & Ree, in press). Drasgow et al (in press) also demonstrated the measurement equivalence of the AFOQT across gender and racial/ethnic subgroups. These results are consistent with analyses of the previous 16 subtest form (Carretta & Ree, 1995b, 1996). The reliabilities for the five composites in the normative sample are: Verbal (.91), Quantitative (92), Academic Aptitude (.94), Pilot (.94), and Navigator/Technical (.95).

Verbal subtests. Verbal Analogies (VA) provides a measure of the ability to reason and determine relationships between words. Word Knowledge (WK) assesses verbal comprehension involving the ability to understand written language through the use of synonyms.

Quantitative subtests. Arithmetic Reasoning (AR) measures the ability to understand arithmetic relations expressed as word problems. Math Knowledge (MK) provides a measure of the ability to use mathematical terms, formulas, and relations.

Spatial subtests. Block Counting (BC) measures spatial ability through the analysis of three-dimensional representations of a set of blocks. Rotated Blocks (RB) assesses the ability to visualize and mentally manipulate objects. Hidden Figures (HF) measures the ability to see a simple figure embedded in a complex drawing.

Aircrew subtests. Instrument Comprehension (IC) assesses the ability to determine the attitude of an aircraft from illustrations of flight instruments. Aviation Information (AI) measures knowledge of general aviation terms, concepts, and principles. General Science (GS) provides a measure of knowledge and understanding of scientific terms, concepts, instruments, and principles.

Perceptual speed subtests. Table Reading (TR) assesses the ability to quickly and accurately extract information from tables.

2.3 Analyses

When conducting a meta-analysis of validities across several studies, it is desirable to correct for artifacts such as sampling error, range restriction, reliability, recording errors, and others that may contribute to variation in outcomes across studies. The extent to which corrections can be made is determined by the availability of data and knowledge about the studies. Hunter and Schmidt (2004) noted that even if all artifacts have been identified and if all known artifacts are controlled, variation in study outcomes due to data errors would still occur. They further noted that in actual meta-analyses attenuation and false variation caused by uncontrolled and unknown artifacts occur in addition to variation caused by bad data. These observations led Schmidt and Hunter (1977) to propose their "75% rule" which serves as a guideline that if in any data set, known and correctable artifacts account for 75% of the variance in study correlations, the remaining 25% of the variance is probably due to uncontrolled artifacts (e.g., study differences in test validity, transcription errors, and typographical errors) and that no substantive variance exists. If variance due to sampling error across the studies accounts for less than 75% of the observed variance, the possibility of moderator variable effects exists.

Whereas previous studies of the predictiveness of AFOQT scores versus performance in non-rated officer specialties have focused on the Academic Aptitude composite (Arth, 1986; Finegold & Rogers, 1985; Hartke & Short, 1988), the current study examined all five composites. Three meta-analyses were performed using observed correlations, correlations corrected for range restriction, and correlations corrected for both range restriction and unreliability of the scores and criterion. Analyses began by examination of the observed correlation between the AFOQT composites and technical training final grade for each officer training course. The meta-analyses of the observed correlations were corrected only for sampling error. The observed correlations then were corrected for range restriction using the multivariate method (Lawley, 1943) and the meta-analyses were repeated. The range-restriction corrected correlations were then corrected for unreliability (Hunter & Schmidt, 2004) of the test scores and training criterion ($r_c = \frac{r_{xy}}{\sqrt{r_{xx}r_{yy}}}$.) and the meta-analyses were repeated. The reliabilities of the measures being correlated affect the correlations. The upper theoretical limit of the correlation between any two measures is the square root of the product of their reliabilities. This third set of

correlations provides a theoretical estimate of the predictiveness of the composites when perfectly reliable measures are available.

3.0 RESULTS

3.1 Observed Correlations

The observed correlations analyses are summarized in Table 2. Ninety percent (63 out of 70) of the observed correlations between the AFOQT composites and average officer technical training grades were statistically significant at or beyond the .05 level. The exceptions occurred for two of the three smallest samples, Combat Controller (n = 16) and Communications Officer Engineering (n = 59). The weighted mean correlations between the AFOQT composites and final technical training course grades ranged from .2614 (Verbal) to .3265 (Academic Aptitude). The proportion of variance accounted for by sampling error was: Verbal (39.62%), Quantitative (75.25%), Academic Aptitude (71.64%), Pilot (61.59%), and Navigator/Technical (72.29%). Of the five composites, only the Quantitative composite met or exceeded Schmidt and Hunter's (1977) 75% threshold. This was evidence of the possible existence of artifacts affecting the variability of the correlations and that the true validity of the Verbal, Academic Aptitude, Pilot, and Navigator/Technical composites was not the same across all occupational specialties for the training courses included in this analysis. Only for the Quantitative composite, where sampling variance accounted for slightly more than 75% of the observed variance around the weighted mean validity, can it be concluded that its validity is the same for all 14 technical training courses and that observed variance in the observed validities is due to artifacts.

Table 2. Correlations between AFOQT Composites and Officer Technical Training Grades: Observed

Air Force Specialty	Course	N	AFOQT Composite V	Q	AA	P	N/T
Combat Control	13D1AB	16	.271	.381	.404	.462*	.484*
Airfield Operations	13M1	251	.266**	.368**	.369**	.382**	.385**
Space & Missile	13S1	1638	.323**	.386**	.426**	.405**	.427**
Space & Missile - Follow-on Course	13S1X	345	.347**	.407**	.435**	.368**	.412**
Intelligence	14N1	1983	.266**	.324**	.353**	.300**	.341**
Weather	15W1	294	.225**	.277**	.311**	.232**	.228**
Aircraft Maintenance	21A1	1430	.243**	.227**	.277**	.274**	.274**
Munitions & Munitions Maintenance - Conventional	21M1C	42	.318*	.377**	.418**	.465**	.398**
Munitions & Munitions Maintenance – Non-Conventional	21M1NC	246	.118*	.238**	.211**	.277**	.288**
Logistics Readiness	21R1	1130	.257**	.243**	294**	.253**	.284**
Security Forces	31P1	599	.295**	.179**	.28**5	.215**	.257**
Communications-Information Systems	33S1	2190	.235**	.250**	.286**	.259**	.282**
Communications Officer Engineering	33S3A	59	.035	.059	.057	.231*	.071
Manpower & Personnel	37F1	319	.216**	.338**	.331**	.297**	.339**
Weighted Mean (All AFSs)		10542	.2614	.2878	.3265	.2965	.3199
95% CI (upper)		10542	.2793	.3053	.3436	.3139	.3371
95% CI (lower)		10542	.2437	.2703	.3095	.2792	.3028

*p ≤ .05; **p ≤ .01

3.2 Range-Restriction Corrected Correlations

The observed correlations were corrected for range retraction (Lawley, 1943) to provide a better statistical estimate of the true relationship between the test scores and training performance. The analyses of the range-restriction corrected correlations are summarized in Table 3. Most correlations increased after correction for range restriction. It should be noted that in a few instances (5 of 70) the correlations decreased in magnitude after correction for range restriction (Combat Control, Communications Officer Engineering). This is unusual, but can occur when the correction leads to a reduction in variance (Levin, 1972). As expected, after correction for range restriction, the weighted mean correlations between the AFOQT composites and final technical training course grades increased for all five composites. The corrected weighted mean correlations ranged from .3222 (Verbal) to .3878 (Academic Aptitude). The proportion of variance accounted for by sampling error also increased for all five composites after correction for range restriction. The values were: Verbal (62.91%), Quantitative (79.27%), Academic Aptitude (78.00%), Pilot (73.97%), and Navigator/Technical (75.61%). After correction, the Quantitative, Academic Aptitude, Navigator/Technical met or exceeded Schmidt and Hunter's (1977) 75% threshold, and the Pilot composite was only slightly below it. Thus, with the exception of the Verbal composite, the predictive validity of the AFOQT composites was the same for all 14 technical training courses and the observed variance in the range-restriction corrected validities can be attributed to artifacts.

Table 3. Correlations between AFOQT Composites and Officer Technical Training Grades: Range-Restriction Corrected

Air Force Specialty	Course	N	AFOQT Composite V	Q	AA	P	N/T
Combat Control	13D1AB	16	.512	.146	.372	.134	.197
Airfield Operations	13M1	251	.286	.389	.393	.407	.410
Space & Missile	13S1	1638	.398	.459	.496	.469	.494
Space & Missile Follow-On Course	13S1X	345	.375	.435	.468	.402	.446
Intelligence	14N1	1983	.338	.388	.420	.365	.407
Weather	15W1	294	.336	.450	.456	.404	.407
Aircraft Maintenance	21A1	1430	.310	.297	.349	.336	.341
Munitions & Munitions Maintenance - Conventional	21M1C	42	.444	.442	.511	.496	.442
Munitions & Munitions Maintenance – Non-Conventional	21M1NC	246	.158	.301	.269	.308	.330
Logistics Readiness	21R1	1130	.307	.297	.348	.296	.332
Security Forces	31P1	599	.355	.255	.347	.282	.324
Communications-Information Systems	33S1	2190	.282	.293	.331	.295	.324
Communications Officer Engineering	33S3A	59	-.005	.160	.095	.333	.188
Manpower & Personnel	37F1	319	.278	.398	.393	.366	.404
Weighted Mean (All AFSs)		10542	.3222	.3499	.3878	.3525	.3796
95% CI (upper)		10542	.3393	.3666	.4040	.3692	.3959
95% CI (lower)		10542	.3051	.3332	.3716	.3358	.3633

Note. No tests for statistical significance were performed for the corrected correlations.

3.3 Range-Restriction and Reliability Corrected Correlations

As previously noted, the upper theoretical limit of the correlation between any two measures is the square root of the product of their reliabilities. Correcting the correlations between the AFOQT composites and technical training scores for unreliability (attenuation) provides a theoretical estimate of the predictiveness of the composites if perfectly reliable measures were available. Reliability estimates for the AFOQT composites were from the Form S normative sample and were based on the Wherry and Gaylord (1943) procedure. The estimates were: Verbal (.91), Quantitative (.92), Academic Aptitude (.94), Pilot (.94), and Navigator/Technical (.95). The reliability of the final technical training grades was estimated to be .80.

The analyses of the correlations after correction for both range-restriction and unreliability of the scores and criterion are summarized in Table 4. Correcting the correlations for both range restriction and unreliability increased their magnitudes above those corrected only for range restriction. The corrected weighted mean correlations between the AFOQT composites and final technical training course grades ranged from .3776 (Verbal) to .4476 (Academic Aptitude). The proportion of variance accounted for by sampling error did not change from the previous analyses where the correlations were corrected for range restriction only. This is because the same reliability estimates were used for all samples. Thus, when the correction for attenuation was applied the validities for each composite were corrected by the same proportion.

Table 4. Correlations between AFOQT Composites and Officer Technical Training Grades: Range-Restriction and Reliability Corrected

Air Force Specialty	Course	N	AFOQT Composite V	Q	AA	P	N/T
Combat Control	13D1AB	16	.600	.170	.428	.154	.225
Airfield Operations	13M1	251	.335	.453	.453	.469	.470
Space & Missile	13S1	1638	.466	.535	.571	.540	.566
Space & Missile Follow-On Course	13S1X	345	.466	.507	.539	.463	.511
Intelligence	14N1	1983	.396	.452	.484	.420	.466
Weather	15W1	294	.393	.524	.525	.465	.466
Aircraft Maintenance	21A1	1430	.363	.346	.402	.387	.391
Munitions & Munitions Maintenance - Conventional	21M1C	42	.520	.515	.589	.571	.507
Munitions & Munitions Maintenance – Non-Conventional	21M1NC	246	.185	.350	.310	.355	.378
Logistics Readiness	21R1	1130	.359	.346	401	.341	.380
Security Forces	31P1	599	.416	.297	.400	.325	.371
Communications-Information Systems	33S1	2190	.330	.341	.381	.340	.371
Communications Officer Engineering	33S3A	59	-.005	.186	.109	.384	.215
Manpower & Personnel	37F1	319	.325	.463	.453	.422	.463
Weighted Mean (All AFSs)		10542	.3776	.4081	.4476	.4071	.4359
95% CI (upper)		10542	.3977	.4277	.4663	.4264	.4546
95% CI (lower)		10542	.3575	.3886	.4289	.3878	.4171

Note. No tests for statistical significance were performed for the corrected correlations.

4.0 DISCUSSION

Results of the meta-analyses of the observed correlations indicated that the predictive validity for four of the five AFOQT composites (Verbal, Academic Aptitude, Pilot, and Navigator/ Technical) was not the same across the non-rated officer training specialties in the current study. Validity generalization was observed for only the Quantitative composite. The lack of generalizability for the observed correlations is consistent with results of a bare bones meta-analysis that examined the generalizability of the validity of the AFOQT Academic Aptitude composite for 47 validity coefficients involving officer technical training grades (Hartke & Short, 1988). Hartke and Short observed that though the validity of the Academic Aptitude composite varied across the officer training specialties, it demonstrated usefulness for nearly all of them. A similar trend was observed in the current analyses. However, the current analyses extended those reported by Hartke and Short, in that the predictive validity of all five composites was examined, not just Academic Aptitude. Though the validity of the composites varied (i.e., was not generalizable) across the 14 officer training specialties, their usefulness was demonstrated for non-rated officer technical training. Ninety percent (63 out 0f 70) of the correlations between the composites and the training criterion were statistically significant.

Hartke and Short (1988) were unable to correct their data for the statistical artifacts of range restriction or unreliability due to a lack of information about the studies in their analyses. The current analyses corrected for both range restriction and unreliability of the scores. After correction for range restriction, four of the five composites demonstrated validity generalization across the training specialties. This means that the true predictive validity of the AFOQT composites (with the exception of Verbal), was consistent across the officer training specialties. The mean validity coefficients for the Quantitative (.3499), Academic Aptitude (.3878), Pilot (.3525), and Navigator/Technical (.3796) composites are the best estimates of the average validity across all officer specialties.

It is interesting to note that though the composites differ in composition (see Table 1), there was little difference in their mean predictive validity. If the Verbal composite (mean weighted validity = .3222) were included despite its lack of generalizability, the range in mean weighted validities was only .0656 (.3878 - .3222 = .0656).

Additional efforts to examine the generalizability of the validity of AFOQT composites should expand to include a broader range of occupational specialties. Expanding the breadth of training specialties would allow the potential moderating effect of occupational similarity to be examined. Occupational subgroups could be defined on the basis of task characteristics of the training specialties. Sorting training specialties into homogeneous subgroups could strengthen validity generalization. Presumably, occupations with similar task characteristics and training also would be more similar in their aptitude requirements. Unfortunately, the current analyses had too few training specialties to allow for division into homogeneous subgroups.

Although the meta-analytic results generally supported the predictiveness of the AFOQT composites across several officer training specialties, results indicate there is reliable variance in training performance not being predicted by the AFOQT composites. Even after correction for range restriction and unreliability, the validities of the composites ranged from .3776 (Verbal) to .4476 (Academic Aptitude). One way to improve predictive validity would be to identify content areas not currently covered by the AFOQT that could account for additional reliable variance in training performance. As previously noted, efforts have begun to identify critical knowledge, skills, abilities, and other characteristics for Air Force officer and technical training programs. The results will be used to guide the identification of constructs to supplement existing AFOQT content.

Finally, as noted earlier, currently there are no AFOQT requirements for non-rated officer training qualification beyond qualifying for an officer commissioning program. The current results suggest that further studies should be conducted to examine the utility of minimum AFOQT qualifying scores for non-technical training specialties, including their effect on training performance and subgroup qualification rates (e.g., adverse impact).

BIBLIOGRAPHY

Arth, T.O. (1986). *Validation of the AFOQT for non-rated officers,* AFHRL-TP-85-50. Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Arth, T. O., & Skinner, J. (1986). *Aptitude selection for Air Force officer non-aircrew jobs.*

Carretta, T. R. (in press). *Predictive validity of the Air Force Officer Qualifying Test for USAF air battle manager training performance,* AFRL-RH-WP-TR-2009-0007. Wright-Patterson AFB, OH: Air Force Research Laboratory, Warfighter Interfaces Division.

Carretta, T. R., & Ree, M. J. (1995a). Air Force Officer Qualifying Test Validity for predicting pilot training performance. *Journal of Business and Psychology, 9,* 379-388.

Carretta, T. R., & Ree, M. J. (1995b). Near identity of cognitive structure in sex and ethnic groups. *Personality and Individual Differences, 19,* 149-155.

Carretta, T. R., & Ree, M. J. (1996). Factor structure of the Air Force Officer Qualifying Test: Analysis and comparison. *Military Psychology, 8,* 29-42.

Carretta, T. R., & Ree, M. J. (2003). Pilot selection methods. In B. H. Kantowitz (Series Ed.) & P. S. Tsang & M. A. Vidulich (Vol. Eds.). *Human factors in transportation: Principles and practices of aviation psychology* (pp. 357-396). Mahwah, NJ: Erlbaum.

Drasgow, F., Nye, C. D., Carretta, T. R., & Ree, M. J. (in press). Factor structure of the Air Force Officer Qualifying Test Form S: Analysis and comparison with previous forms. *Military Psychology*.

Finegold, L., & Rogers, D. (1985). *Relationship between Air Force Officer Qualifying Test scores and success in air weapons controller training,* AFHRL-TR-85-13.Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division

Hartke, D. D., & Short. L. O. (1988). *Validity of the academic aptitude composite of the Air Force Officer Qualifying Test (AFOQT),* AFHRL-TP-87-61. Brooks AFB,

TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis.* Thousand Oaks, CA: Sage.

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh, Section A, 62, Part 1,* 28-30.

Levin, J. (1972). The occurrence of an increase in correlation by range restriction. *Psychometrika, 37,* 93-97.

Olea, M. M., & Ree, M. J. (1994). Predicting pilot and navigator criteria: Not much more than *g*. *Journal of Applied Psychology, 79,* 845-849.

Roberts, H. E., & Skinner, J. (1996). Gender and racial equity of the Air Force Officer Qualifying Test in officer training school selection decisions. *Military Psychology, 8,* 95-113.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62,* 529-540.

United States Air Force (2006). *Officer training school (OTS) and enlisted commissioning programs (ECP), Air Force Instruction 36-2013.* Washington, DC: Department of the Air Force.

Wherry, R. J., & Gaylord, R. H. (1943). The concept of test and item reliability in relation to factor pattern. *Psychometrika, 8,* 247-264.